Towards Secure Mobile Multiagent Based Electronic Marketplace Systems

Klaus Fischer

German Research Center for Artificial Intelligence (DFKI GmbH)
66123 Saarbrücken, Germany

Dieter Hutter

German Research Center for Artificial Intelligence (DFKI GmbH)
66123 Saarbrücken, Germany

Matthias Klusch

German Research Center for Artificial Intelligence (DFKI GmbH)
66123 Saarbrücken, Germany

Werner Stephan

German Research Center for Artificial Intelligence (DFKI GmbH)
66123 Saarbrücken, Germany

Abstract

In this paper we aim at a generic methodology to validate, assess, and construct mobile multi-agent systems in the domain of *virtual market places*. It rests upon research carried out in the past to develop innovative engineering solutions for architectures based on the paradigm of *mobile multi-agents*. This methodology will cover requirement analysis including imposed security issues, system design, and implementation of such systems. Virtual market places of different types will serve as application scenarios, where *security issues* are addressed in all phases of the development process.

Rather than treating the various technical aspects in isolation, we want to integrate research and development on economic models of market places, (formal) security requirements, functional architectures of (mobile) agents and societies of agents, security functions, and the implementation of the architectural design and abstract security solutions on existing platforms.

Although commenced by looking at concrete (market place) scenarios, the ultimate goal is to come up with a generic methodology that provides analytic techniques, techniques for (formal) descriptions, and design patterns that are useful for a large variety of (market place) scenarios and other application areas.

---

1 Project Goals

The success of the Internet and the World Wide Web (WWW) has deeply influenced our every day lives as well as our business and commercial structure. Forrester Research estimates that Internet-based business-to-business transactions will grow from 8 billion dollars in 1999 to at least 327 billion dollars in 2002.

Agent technologies and multiagent systems will play a major part in the further development of WWW-based applications: virtual market places with customer agents and seller agents, chat rooms and avatars, personal assistant agents as well as non-benevolent agents designed to attack a site, are just some of the many applications.

Successful attacks, most recently for example the intrusion into the local network of Microsoft, demonstrate the danger globally operating companies face. Also the danger of intrusion into military head quarters is well-known. Forrester Research reports that despite an expected 300 percent increase in spending on information technology security over the next four years, U.S. companies will still be left almost as vulnerable to security breaches as they are today. Agent technologies increase the possibilities of malicious practice: e.g it is technological feasible to design a swarm of attack agents, which is able to reproduce itself on any server it gets access to, and which can carry viruses or Trojan Horses, or just extract classified information.

Research in security has been focused on the solutions of individual security problems. For example, various solutions have been proposed to define and verify secure key-exchange protocols or to translate a given trust model into a formal security policy. However, only little research has been carried out to integrate these individual techniques into a global security methodology for basic agent technologies and multiagent systems.

In this paper we investigate fundamental security threats in the design of multiagent systems within virtual market places. These threats can be classified whether they are: (i) inherent to the application scenario to be implemented, (ii) inherent in the multiagent system level design, or (iii) a consequence of the design of an individual agent, or (iv) a result of using mobile computing. We therefore investigate into how the design of the application, the design of the agent society (as well as the individual agents), and the selection of the computation paradigm influences the characteristics of the security threats and how security measures can be combined to an all-embracing security infrastructure. As a result, we aim at providing a specific

methodology that uses the multiagent system paradigm to realize secure applications. Accordingly, this methodology (and the project work) is organized around three levels: the *application architecture*, the *system architecture*, and the *computational architecture*.

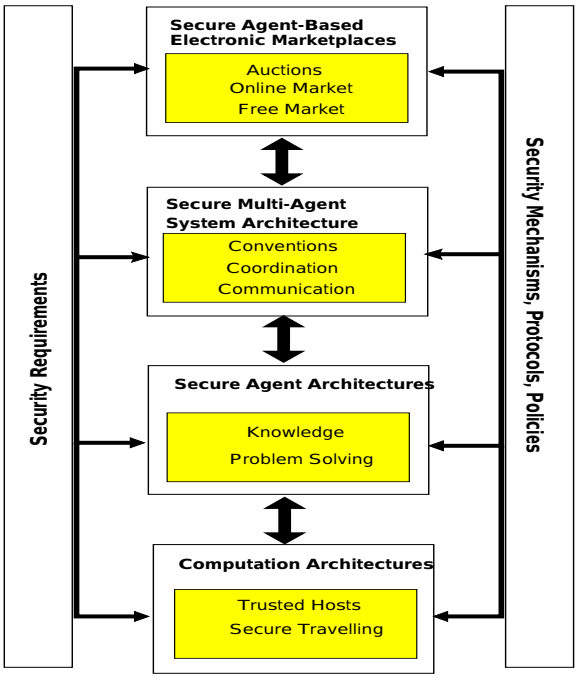

Fig. 1. Layers of Security

We use concrete instances of virtual market places to illustrate our approach. This application domain is interesting for multiagent system research in its own right. However, our multiagent system research will concentrate on a compositional design methodology for multiagent system to investigate how top-down requirements and bottom-up behavior of multiagent system can be brought together.

2 Motivating Example

SEMAS will investigate several settings for virtual market places, i.e. markets based on auctions and free negotiation. In the following we use the model of a virtual mall to illustrate the research topics that are relevant for SEMAS in the context of two concrete example scenarios as shown in Figure 2.

These scenarios, mobile comparison shopping and an auction house at a virtual mall, will serve as *use cases for the research and development of the SEMAS methodology*. Both scenarios, and their corresponding security issues, are described in more detail in the following sections.

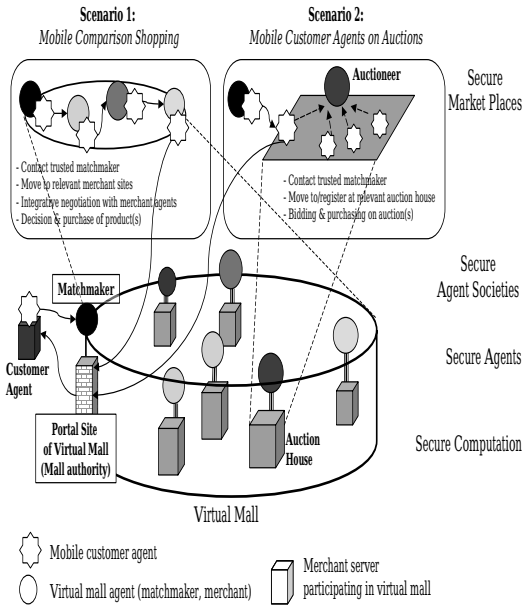

Fig. 2. Example scenarios for mobile agents at virtual mall

2.1 Virtual Mall: Mobile Comparison Shopping

This scenario deals with integrative negotiation in the course of mobile comparison shopping agent within a virtual mall. A mobile customer agent first contacts a matchmaker agent and, based on the result, then visits selected merchants' sites within the mall that can potentially contribute to its search for the best price and other relevant attributes for the desired product the user wants to purchase. For this purpose, the agent negotiates with the merchant agents based on multi-attribute theory [10].

The multi-attribute theory provides a way of representing and calculating the utilities of outcomes of actions by decomposing them into utilities of the value-relevant factors that make up the outcomes of recommendation actions. Recommendation is determined by defeasible decision making on the basis of the values assigned to different factors and the magnitude of these values, respectively. In addition, this theory offers methods for composing utility measures and the construction of libraries of standard forms for utility functions.

The customer agent recommends product items to the user based on an evaluation of multiple attributes such as price and quality of the desired products, delivery times and costs, return policies, promotions and gift services as well as customer support and reputation of respective merchants. The underlying negotiation between user and merchant agents analyzes the decision problem of what and who to buy from in terms of transaction, or more qualitatively, through attribute constraints. It solves this decision problem by applying finite-domain constraint satisfaction techniques [18].

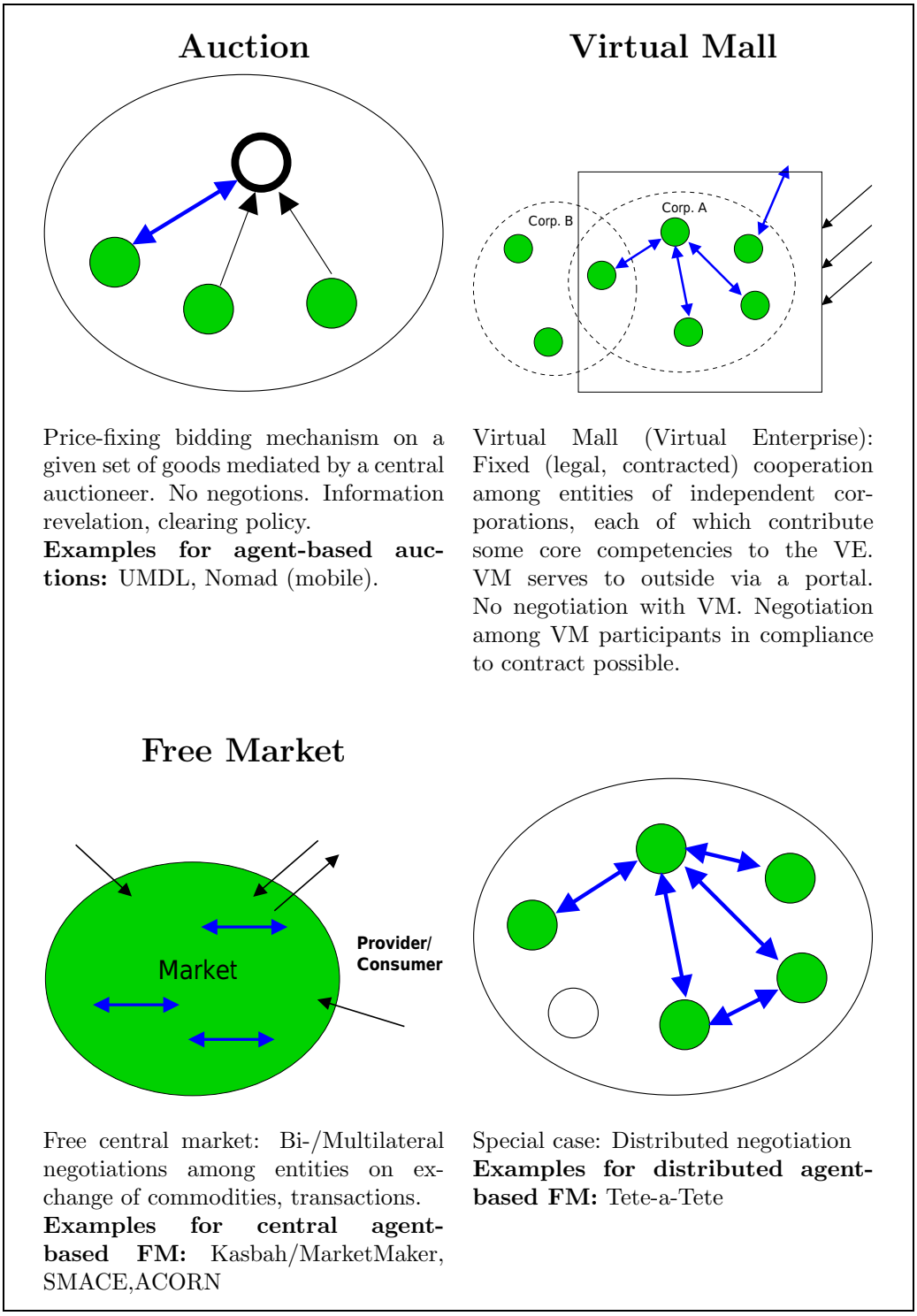

Fig. 3. Different Kinds of Electronic Market Places

A portal makes it easier for the customer to deal with large malls as only one address and corresponding public key has to be known. Empowered by the matchmaker agent service of the portal site, the mobile customer agent visits merchants within the mall that can contribute without the necessity for

the user to be connected to the mall at all times. It returns with the result when the user is online again. Since the user trusts his mobile agent it does not need to interact with the user during the negotiations and visits.

Like intermediaries in the physical economy, intelligent middle-agents such as matchmaker and broker agents can be considered as electronic intermediaries in the digital economy [1]. These agents provide means for coordinating activities among agent providers and requesters of services (information, goods, or expertise) in the Internet [3]. Their main task is to locate and connect the ultimate service providers with the ultimate requesters in open environments, that is to appropriately cope with the connection problem.

The purpose of the mall is to let agents travel from server to server while collecting intermediate results. Although all the nodes are known beforehand, the agent is free to decide how many nodes in the mall to visit, and in what order. This means that at any place the agent can determine, based on its current state, its next hop or hops. The intermediate results are the offers of the selected merchants that are returned to the agent. The agent evaluates these offers concurrently and decides at the end of the visits with whom to make the deal.

2.1.1 Security Issues of Mobile Comparison Shopping

For the comparison shopping scenario outlined above we shall now describe the security engineering aspects of the SEMAS approach in somewhat more detail. For each of the (abstraction) levels we indicate the relevant security issues, modeling techniques, and design steps.

At the global level of the application architecture, the basic organization of the market place scenario is defined without looking at specific technical solutions at this point of time. In the corresponding phase of security engineering a global security policy is developed by identifying threats and defining security objectives to counteract these threats. In a rigorous approach both, the abstract application architecture and the desired security objectives have to be formalized. The formalization of the Digital Signature Scenario, which was done as part of the VSE-II project, is an example of a "formal security policy". The main challenge comes from formalizing security objectives at this abstract level.

Since the application architecture will be defined without referring to concepts of a particular technology, such as special agent architectures, we will use the general terminology for the assessment of IT-technology, like the Information Technology Security Evluation Criteria ITSEC, [14], and the Common Criteria, CC, [6] as a starting point. Where necessary these concepts will be refined or extended [1]. Risk analysis in this setting proceeds by identifying *subjects* acting in certain roles and classes of *objects* owned (used, manipulated)

[1] This may be necessary, as the prevailing terminology of these criteria stems from the (security) analysis of operating systems

or exchanged by these subjects. Depending on the particular application, subjects can be persons, organizations, and computer processes. Of course, in our context subjects will most often be realized by agents. In the comparison shopping scenario, for instance, the most important roles are that of merchants, customers, and matchmakers. Classes of objects include products, offers, orders, or agents.

A *separation of concerns* is provided by a classification of security issues. In our scenario one could distinguish between *trust*, *confidentiality*, *integrity*, *non-repudiation*, and *anonymity*. This classification supports a rigorous analysis and is useful for subsequent technical solutions. For example, it seems to be necessary that one distinguishes between the confidentiality of messages exchanged between subjects and the trust that the recipient will handle the information according to certain rules. On the other hand there might be potentially conflicting issues such as the necessity to obtain trust in (the behavior of) certain subjects and the request for anonymity of these subjects.

In certain situations it is necessary that a subject can *trust* another subject with respect to certain aspects of its behavior. This holds in particular for electronic marketplaces as constituents of the Internet economy. As mentioned above, middle-agents may act as intermediaries on such marektplaces. It is an essential requirement of any middle-agent to behave in a trustworthy manner to its clients [20,13]. In this sense, a middle-agent acts as a trusted intermediary among the agents of the considered agent society according to given trust policies of individual agents and the whole society.

Both internal data and the computational processes of mediation executed by the middle-agent should be robust against external manipulation or attacks of malicious agents, systems, or users. On the other hand, clients of a middle-agent should have no incentive to misuse any information which has been revealed by the middle-agent during mediation. In this respect most common trust actions are authorization and verification of credentials of all parties which are involved in the mediation.

Actions have to take different trust relationships into account which may exist among the customer agents, middle-agents, and the provider or merchant agents within one or more connected agent societies. Can a contacted middle-agent, for example, be trusted by its clients to not resell (parts of) private profile information, copyrighted data, and so forth, to other middle-agents or clients it collaborates with? Does it also hold across multiagent systems borders? Customer agents also might want to verify certain facts about relevant merchant agents before contracting them.

In summary, there is an obvious need for all agents involved in the mediation process to use an appropriate trust model to analyze and assess the risks of and methods to prevent and counteract attacks against their data and knowledge. Models, methods and techniques supporting the establishment and management of mutual trust between a middle-agent and its clients in an open environment include:

- the use of expressive trust establishment certificates useful for, e.g., binding agent identity to public key infrastructures such as IETF's SPKI [5,17], and other standard security mechanisms and protocols, the use of mechanisms to bind agent names to their human deployers, so that the human would bear responsibility in case his/her agent misbehaves.
- the formal specification of agent trust policies, and
- the respective update, propagation, and transitive merge of trust matrices to calculate an overall trust relationship that accounts for the trust values in each and every individual trust relationship which is relevant to the mediation process [12]. An option is the application of distributed history-based reputation mechanisms to agent societies [21].

In the context of a virtual mall the following issues of trust may be identified. The customer has to be sure that the subject she addresses as a matchmaker really represents the virtual mall she wants to visit. Moreover, the customer will expect from the matchmaker that it provides her with an exhaustive collection of merchant's locations, that are relevant for her search and she can trust in, and that the matchmaker handles the information given to it by the customer according to certain rules. Vice versa, the matchmaker has to identify the subject, addressing it as an approved customer which among other criteria has to be examined according to her credit standing. Note that this does not necessarily mean that the matchmaker has to know about the final identity of the customer.

Confidentiality means that unauthorized subjects should not be able to gain knowledge about messages (objects) exchanged between other subjects or data owned by other subjects. In a scenario such as this, one has to fix rules with respect to the right to know certain information or, on the contrary, with respect to keep information secret. In our example scenario the offers made by a particular merchant to a customer should not become known to other merchants and customers. This also holds for third parties that are part of the virtual mall and for subjects outside that are able to perceive the communication process going on. In contrast to this, it is a matter of trust how information is treated by authorized subjects. There are various formal approaches to model confidentiality. In interference/noninterference techniques as they have been used, for example, in the VSE security models for chip cards, there is a distinction between allowed/disallowed flow of information. The FM-DIN security model, [11], used an explicit theory of *knowledge* (acquisition). Saying that an offer should not be compromised does not simply mean that the data object sent to the customer should not become known to some "attacker". Rather, it should be impossible to *derive* the price mentioned in the offer from the information accessible to an "attacker".

Unauthorized subjects should not be able to manipulate objects exchanged or owned by other subjects thereby violating their *integrity*. For example, merchants should not be able to manipulate the offers collected by a customer

moving around in the mall. As in the case of confidentiality, allowed and disallowed changes have to be modeled in an appropriate way. While a customer might be given the right to look at the recommendations given by the matchmaker to another customer it should be impossible for him to manipulate this information. Techniques to formalize integrity are often similar to those used for confidentiality.

At least in some situations it should by impossible for a subject to reveal the identity of another subject. So (perhaps) a customer should be able to walk around in the mall and collect offers without being identified. Nevertheless, the matchmaker has to be able to check whether the customer is allowed to visit the mall. As opposed to confidentiality, *anonymity* does not deal with the content data objects but with the source of a communication, although at a more technical level information about the source and destination of a message can be part of the message itself. Note that anonymity does not mean that the identity of a subject is hidden to all other subjects for all time. So for example, there will be a trusted authority that knows about the identity of all customers. Moreover, it seems sensible that in later stages the identity of a customer becomes known to a particular merchant.

Non-repudiation is concerned with the legal relevance of communications. To make a merchants offer binding the customer has to be in a position where he can prove that the message (containing the offer) was willingly sent by the merchant.

The system design process starts with a description of technical solutions at an abstract level. In our case this means that the virtual mall scenario is mapped to a software architecture. Without going into the details of an efficient implementation (programming languages, platforms, communication) the basic algorithms determining the behavior of agents making up the virtual mall scenario are designed and specified. This includes those parts that are critical with respect to the security issues identified in the first phase.

The abstract security policy has to be realized by what is sometimes called *security functions and mechanisms*. For our scenario the following functions are relevant:

- Trust can be realized by certificates.
- Confidentiality can be achieved by access control and encryption.
- Integrity can be achieved by access control and can be checked by using hash-functions.
- Anonymity can be achieved by using certified pseudonyms.
- Non-repudiability can be achieved by using digital signatures.

2.2 Virtual Mall: Mobile Customer Agents and Auctions

This scenario concerns participation of mobile customer agents in one or more auctions at one or more auction houses in a virtual mall. Each auction house is

owned by one or more merchants in the mall and may run one or more auctions for a given type of commodity or task, that is a set of items each of which multiple customers can concurrently bid for. As in the mobile comparison shopping scenario a mobile customer agent first contacts a matchmaker agent at the portal site of the virtual mall to obtain information on the relevant auctions and auction houses in the mall. Based on these recommendations the agent moves to selected relevant auction servers sequentially to participate in respective auctions. It may also reside on a safe server near by an auction house to track ongoing auctions. Before we discuss the benefits of using mobile agents on auctions and their security issues, let us briefly recall the notion of an auction.

Auction theory [19] analyzes protocols and agents' strategies in auctions. An auction is a price-fixing mechanism of an auction house in which negotiation is subject to a very strict coordination process. It consists of an auctioneer who wants to mediate the exchange of given items between buyers and vendors for sale at the highest possible price, and potential bidders who want to buy them at a lowest possible price. Asynchronous bidding mechanisms are mostly based on open-outcry with price changes or sealed bids with periodic partial revelation. Any auction is a sequence of auction rounds and concerns auctioning of goods or tasks. The private value of an item depends only on the agent's own preferences; its common value corresponds to the agent's value of the item determined by others' values; its correlated value depends partly on the agent's own preferences and partly on others' values for it.

Online auctions appear to be unnecessarily hostile to customers due to the winner's curse and offer no long-term benefits to merchants. If bidders have reasonable information about the worth of the item, then the average of all the guesses is likely to be correct. However, the winner offers the bid furthest from the actual value, thus, pays more for the item than its value, so any auction is basically a win-lose game.

Any auction may be classified along three dimensions of (1) bidding rules including, for example, bid format, and many-to-one or many-to-many participation, (2) clearing policy such as pricing, clear schedule and closing, and (3) information revelation policy including, for example, price quotes, quote schedule, etc.

Prominent protocols for auctions we will consider in the scenario include the

- first-price, open-cry, so-called *English auction*. The bidders successively raise a bid for an item until one bidder remains. The winner is the last bidder remaining at the price of the second-highest bidder. The dominant strategy for consumers here is to bid up to their true (private), maximum value, then drop out.
- descending price, open-cry, so-called *Dutch auction* that guarantees the auctioneer the purchase of items at highest possible price. The rules are that

the auctioneer calls out a descending price for an item and the bidder call out a bid. The winner is the first bidder to call out at a price bid. Optimal strategy is to bid just below private value of item.

- first-price, sealed-bid auction where each bidder submits one bid in ignorance of all other bids. The highest bidder wins and pays the amount he bids. This has the potential to force buyers and seller into price wars since the sealed bid of any bidder depend on what s/he believes of all other opponents bids.
- second-price, sealed-bid, so-called *Vickrey auction* where the winning bidder pays the price of only the second highest bid [16].

Under the assumption of subjective private value, all four basic auction types listed above can be shown to yield the same expected price and revenue to the seller when bidders are not risk-averse but risk neutral and symmetric (means they use the same measurements to estimate their valuations). This implies that the auction choice is not as crucial because each format yields on average the same payoff. But revenue equivalence does not hold true under common value assumption (when bidders have similar evaluations). Auction types for multiple (identical or heterogeneous) items for sale are, for example, the discriminatory, the double, and the matrix auction. Further auction-based mechanisms are discussed in, for example, [4].

What are the main benefits for customers of using mobile agents to participate in auctions in a virtual mall? Mobile customer agents roaming the network of the virtual mall may monitor events and track bids in multiple auctions to help their customers for conditioned bidding in time and best deals. The benefits of using mobile agents which move near or on the auction house server include, amongst others,

- avoidance of network lag to get information and bid more timely,
- continuation of bidding even when the user is temporarily disconnected, and
- avoidance of large sizes and frequent downloading of information, thereby saving bandwidth and time for bidding decisions.

2.2.1 Security Issues of Mobile Agents on Auctions

The security issues of trust, confidentiality, and integrity in this scenario are closely related to the types of threats to auctions of any of the types mentioned above. These include, amongst others, vulnerability of auctions to (1) *bidder collusions*, (2) *shills*, (3) *lying auctioneers*, and (4) *undesirable private information revelation*.

A serious threat to auctioning from the customers' perspective concerns so-called shills, which means that special agents are planted by the auctioneer to manipulate the valuation of the auctioned good by raising bids to stimulate the market. However, shills are only a problem in non-private value settings for English and all-pay auctions. Vickrey, first-price sealed-bid, and Dutch

auctions are not vulnerable to this threat. In general, classic analysis of auctions ignore the possibility of shills violating the trust in auctioneers and the integrity of offers.

The same goes with lying auctioneers. For example, in Vickrey auctions, a lying auctioneer can overstate the second highest bid to the winning bidder in order to increase its own revenue. This implies the need of bid verification mechanisms, e.g. via cryptographic signatures. Another possibility would be the use of third-party auction bots which will reveal the (real) highest bid to the seller after closing of the auction. In public auctions auctioneers have no incentives to lie to the bidders. However, the auctioneer might try to refuse to sell after the auction has ended.

Another threat concerns confidentiality violated via undesirable revelation of private information by both the auctioneer or customer agents. There are observable problems with subcontractors of bidders whose strategic marginal cost information is revealed in Vickrey and English auction since truthful bidding is a dominant strategy in such types of auctions. In general, bidding is often the result of correct predictions about the behavior of others and sometimes that means guessing the extent of someone else's information correctly. A mobile customer agent may attack other competing customer agents to gather such information concurrently to support the bidding of its customer. On the other hand an auctioneer could sell or reveal private information on bids and preferences of past and current bidders to competing agents for misuse.

Regarding all of the above types of threats the main security issues for mobile customer agents are trust in, and integrity of, the auction house as well as confidentiality of both, auction house and customer agents.

From the perspective of an auction house a different threat concerns coalition formation among customer agents during an auction which agree to not outbid one another but distribute the purchased items among themselves privately. This is an issue of trust of an auctioneer in the customer agents which are currently participating in the auction. Both issues, shills and bidder collusions or coalitions on auctions are considered illegal but hard to detect; thus, mechanisms to reduce the agents' incentive for both types of actions a-priori have to be embedded into the negotiation protocol explicitly or indirectly as part of the protocol's theoretical features. Please note that the formation of coalitions among mobile customer agents in a virtual mall is not prohibited: a coalition as a whole is allowed to participate in an auction where a representative is chosen as the bidder.

3 A Methodology

We aim at the development of innovative engineering solutions for architectures based on mobile multi-agent systems in an e-commerce setting. E-commerce can be defined as the sum of all activities that are directly con-

cerned with the trade of goods and services on the Internet. It is thus part of the more general electronic business communication of so-called e-business activities. E-Commerce can be sub-divided into separate market segments according to the direction of the flow of trade between business areas or producers and consumers of goods.

Implementing e-commerce applications, results in large and complex systems which cannot be developed without anticipating and targeting possible vulnerabilities. However, to determine, analyze, and assess these vulnerabilities requires a thorough knowledge of how the individual components interact with each other. The complex and distributed nature of such systems raises the question of how to combine security measures, targeted at security issues of individual components to an overall security infrastructure. Moreover, trustworthy systems may not presuppose that all components are trustworthy. Vulnerabilities of particular components may be circumvented if the environment using these components takes care of their weaknesses.

In order to construct trustworthy large-scale systems in the domain of mobile multi-agent systems, a general methodology has to map (decompose) the notion of trustworthiness of an overall system to specific requirements, or trustworthiness of its individual components. Rather than improving particular solutions for small-scale security problems, the problem is to find an appropriate decomposition of the overall system with respect to security, allowing us to (re-)use existing approaches for small-scale problems. We propose three major levels for mobile multi-agent systems: the application architecture, the systems design and the computational level. In the following, we will sketch the issues which arise at each individual level, and between successive levels (see Figure 1).

3.0.2 Application Architecture — Secure Market Places:

The application architecture is concerned with the overall design and organization of the market place scenarios under consideration. This is done without looking at technical solutions.

The multi-agent research at this level is concentrated on the design and analysis of innovative models of virtual market places. Rather than coming up with new economic models for virtual market places, the idea is to adapt existing models to the case of multi-agent systems. From todays perspective auction houses, virtual malls, and free markets are the basic models that form the starting point for this thread of research. Concrete instances of these models have to be designed and will form the basis for further investigations. The main focus of multiagent system research at this level is to investigate the interaction of security requirements with the general model specification of virtual market places.

In the corresponding phase of security engineering, potential threats are identified and security objectives are defined to counteract these threats. In a rigorous approach both, the abstract application architecture, and the desired

security objectives, are formalized.

At the application architecture level, we investigate on how security requirements of the different types of virtual markets can be formalized, and later-on guaranteed (verified) depending on which security mechanisms and protocols are provided by the underlying systems architecture level. Abstract specification of negotiation and interaction models are developed, according to the underlying economic models. Furthermore, we we investigate in how changes of these underlying economic models influence the definition and verification of security objectives.

For example, an appropriate management of trust has to be defined for each application scenario which guides the information flow within the agent society. Formal security policies constitute the technical means to formalize such a management of trust. We will, for instance, be concerned with the development and instantiation of non-interference policies for non-deterministic systems to allow for the verification (validation) of the security objectives of the particular multi-agent systems under consideration.

3.0.3 System Architecture — Secure Agents and Agent Societies

The system architecture is concerned with the technical solution that realizes the intended virtual market places based on a society of agents. This level is concerned with both the extention and refinement of existing multi-agent systems to suit the design of virtual market places, and the further development of the underlying architectures for specifying individual agents.

The design of multi-agent societies have to meet the requirements of the models specified at the application architecture level. Negotiation mechanisms among individual agents are needed in these agent societies. According to [7] negotiation research can be considered to deal with three broad topics: negotiation protocols, negotiation objects, and agents' decision making models. While it is clear that one has to deal with all three issues we want to put a focus on negotiation protocols and decision making models.

We investigate the interaction between the architecture of a single agent and that of the emerging multi-agent system particularly with respect to security requirements. The question arises as to which types of secure architectures for agent societies are possible with respect to knowledge, problem-solving, data storage, communication, social conventions, and protocols. We aim at a formal definition of instantiations of these architectures to allow for a gradual verification and evaluation. The agent architecture InteRRaP-R [9,8], which was developed at the DFKI, is the starting point for this research. However, it is necessary to redesign InteRRaP-R to meet the security issues of the agent architecture. Therefore, we investigate in how far the reasoning procedures in InteRRaP-R can be adapted to reason about basic security mechanisms as well. Furthermore, the reasoning procedures are extended to allow meta-reasoning to give the agent the ability to find out whether executed actions have the intended effects.

At the level of individual agents, security issues can be classified as: the correct functional behavior, the secure traveling through the network, and the secure communication with other agents. In the past, a variety of formal methods have been developed to prove that a program (an agent) behaves according to its (formal) specification. To guarantee secure traveling of agents, agents have to authorize themselves to the host. After a successful authorization, an access policy defines the resources an agent may use on a foreign host. The question arises how to define and implement such access policies allowing, on the one hand, the agent to access all necessary data, and on the other hand, the host to charge the authority of the agent for the used resources. If agents are endangered when traveling on the net, the use of multiple instances of individual agents might be an appropriate approach to provide more robust system behavior.

Communications between agents have to be secured with the help of cryptographic (security) protocols. The question arises as of how existing security protocols can be used and how this selection will effect the construction of a secure agent architecture (or system). For instance, the threat that a host may attack an agent results in the restriction that agents must not possess a secret key. Therefore, agents are not able to sign documents unless they make use of a trusted authority or migrate to a trusted host. Static information of the agent may be used to formulate the digital signature. In this context, the possibility of splitting information across different agents [15] might be an interesting concept which could lead to innovative solutions.

At the level of agent societies the question arises as to what kind of architectures and implementations of secure agents and mobile multi-agent systems are compatible with the security requirements at the application level. Which types of secure agents are compatible with which types of secure market places, and how, and in what way, this compatibility is gradually verifiable, or amenable to evaluation? Conversely, the following question arises: what are the provable effects in the theory and implementation of which kinds of security mechanisms at the level of the agent – and a correspondingly secure, generic architecture of an agent (or an agent system) on the economic model of the respective market places? For example, could the restrictions obtained this way maintain theoretical convergence?

It is well know from other applications that the relation between the application architecture and the abstract system design is not a simple refinement. For example, in order to use encryption, complex (cryptographic) protocols for the exchange of keys have to be introduced that require an analysis of their own. The management of certificates raises new integrity problems. Basically, the whole analysis has to be redone at a new level. The situation is even worse if mobile agents are considered. Implicit assumptions, like the integrity of customers themselves, require complex mechanisms if these are realized by agents that migrate to possible hostile computation platforms.

3.0.4 Computational Architecture

The implementation architecture is concerned with the provision of a secure environment for distributed computation. Security requirements at this level are concerned, for instance, with threats that an agent attacks a host or that a host accesses an agent's secrets. Threats to a host may endanger the integrity of a host. While proper access control mechanisms (like Java Virtual Machine, JVM) can be used to ensure that agents do not change the intentional behavior of the hosts, it is still an open problem how to prevent a host from attacking an agent. In principle, an agent cannot verify the correct execution of its code.

For example, a host may either alter, read, or delete confidential information of the agent or it may slow down or even refuse the execution of an agent's program. Methods are needed to protect confidential information of an individual agent when resident on a host as well as when migrating from one host to another. The agent information can be divided into different types.

The agent's program code and other static information can be protected by digitally signing it, thus **after-the-fact** detection of tampering is possible. However, it is impossible to **prevent** agent tampering unless trusted hardware is available at the foreign host. Since the agent's code has to run on a foreign host, it is also impossible to keep its code private. In order to keep its code secret, an agent has to resort to using a trusted host.

Since an agent's foremost task is to gather information, the privacy of this information is usually of greater concern than the privacy of the agent's code. There are two modes of information gathering [2]. In a *stateless* mode, an agent sends the collected information home to its authority. Thus, the foreign host providing the information can also be used to take care of encryption and of the delivery of the information to the authority. In a *stateful* mode, the gained information is in some way attached to the agent and it carries it along its way when migrating from one host to another. Again the host providing the information may encrypt the data with the help of the public key of the agents authority preventing the disclosure of the information to other hosts. However, in order to make confidential use of the collected data during its itinerary, the agent has again to resort to a trusted (e.g. tamper-resistant) host. Thus, the agent's itinerary will be influenced by security aspects. An agent has to deliberately migrate to trusted hardware in order to process its collected data.

4 Conclusion

In this paper we descirbed a generic methodology to validate, assess, and construct mobile mult-agent systems in the domain of virtual marketplaces. This methodology distinguishes three levels of abstraction on which design and security aspects can be investigated: the application architecture, the system architecture, and the computational architecture. The main contribution of this article and the work done so far was to specify this general framework.

As it was outlined in the paper there are results to deal with specific security problems in each of the abstraction levels. The main aim of our future work is therefore to investigate an integrative model which can deal with the interaction of security mechanisms that are specified and implemented at different levels.

References

[1] A. Barua, A. Whinston, and F. Yin. Value and productivity in the internet economy. *IEEE Computer*, May 2000.

[2] D. Chess, B. Grosof, C. Harrison, D. Levine, C. Parris, and G. Tsudik. Itinerant agents for mobile computing. Technical report, IBM T.J. Watson Research Center, N.Y., October 1995.

[3] K. Decker, K. Sycara, and M. Williamson. Middle-agents for the internet. In *IJCAI-97 International Joint Conference on Artificial Intelligence, Nagoya, Japan*, 1997.

[4] K. Fischer, C. Ruß, and G. Vierke. Decision Theory and Coordination in Multiagent Systems. Research Report RR-98-02, DFKI, 1998.

[5] A. Herzberg, Y. Mass, J. Mihaeli, D. Naor, and Y. Ravid. Access control meets public key infrastructure, or: Assigning roles to strangers. In *IEEE Symposium on Security and Privacy*, May 2000.

[6] ISO/IEC International Standard. *Common Criteria for Information Technology Security Evaluation (CC), Version 2.1, ISO IS 15408*, 2000. Available from `http: //csrc.ncsl.nist.gov/nistpubs/cc/`.

[7] N. R. Jennings, S. Parsons, C. Sierra, and P. Faratin. Automated negotiation. In *Proc. 5th Int. Conf. on the Practical Application of Intelligent Agents and Multi- Agent Systems (PAAM-2000)*, pages 23–30, Manchester, UK, 2000.

[8] C. G. Jung and K. Fischer. A Layered Agent Calculus with Concurrent, Continuous Processes. In *Intelligent Agents IV*, volume 1365 of *Lecture Notes in Artificial Intelligence*, pages 245–258. Springer, 1998.

[9] C. G. Jung and K. Fischer. Methodological comparison of agent models. Technical Report RR-98-1, DFKI GmbH, Saarbrücken, Germany, 1998.

[10] R. L. Keeney and H. Raiffa. *Decisions With Multiple Objectives*. John Wiley and Sons, New York, 1976.

[11] B. Langenstein, M. Ullmann, and R. Vogt. The use of formal methods for trusted digital signature devices. In *Proc. 13th Intern. FLAIRS Conf.* AAAI Press, 2000.

[12] D. Manchala. E-commerce trust metrics and models. *IEEE Internet Computing*, 4(2), March/April 2000.

[13] Y. Mass and O. Shehory. Distributed trust in open multi-agent systems. In *Autonomous Agents 2000 Workshop on Deception, Fraud and Trust in Agent Societies, June*, 2000.

[14] Office for Official Publications of the European Communities. *Information Technology Security Evaluation Criteria (ITSEC), Version 1.2*, 1991.

[15] V. Roth. Mutual protection of co-operating agents. In J. Vitek and C. Jensen, editors, *Secure Internet Programming: Security Issues for Mobile and Distributed Objects*. Springer, LNCS 1603, 1999.

[16] T. Sandholm. Limitations of the vickrey auction in computational multiagent systems. In *Proc. of Intl. conference on multiagent systems (ICMAS)*, 1996.

[17] SPKI Simple PublicKey Infrastructure. ftp://ftp.ietf.org/internet-drafts/draft-ietf-spki-cert-theory-02.txt.

[18] E. Tsang. *Foundations of constraint satisfaction*. Computation in cognitive science. Academic Press, 1993.

[19] E. Wolfstetter. Auctions: An introduction. *Journal on Economic Surveys*, 10(4), 1996.

[20] H. Wong and K. Sycara. Adding security and trust to multi-agent systems. In *Autonomous Agents 1999 Workshop on Deception, Fraud, and Trust in Agent Societies*, May 1999.

[21] B. Yu and M. Singh. A social mechanism of reputation management in electronic communities. In M. Klusch and L. Kerschberg, editors, *CIA-2000 Workshop on Cooperative Information Agents*, volume 1860 of *LNAI*. Springer, 2000.